

přednáška 11:
test střední hodnoty průměru při **NEznámém
rozptylu**

a) test $\mu_1 = \text{const}$

b) test $\mu_1 = \mu_2$

Literatura v IS:

Přednáška by měla být přepsána během pondělka 10. května 2021 do textu Ma0008.

Dosud probrané testy statistické:

B) test střední hodnoty binomického rozdělení

C) Test střední hodnoty průměru při známém rozptylu

A nyní se zaměříme na

D) Test střední hodnoty průměru při neznámém rozptylu

Test střední hodnoty průměru při neznámém rozptylu:

V reálných datech rozptyl σ^2 většinou neznáme, tj. při pravděpodobnostním či statistickém popisu jej musíme odhadnout

- Odhad rozptylu budeme značit est σ^2
(z anglického „estimate“ = odhad)

První adept na odhad neznámého rozptylu na základě měření:

$$\text{est } \sigma^2 = s^2 = \left(\frac{1}{N} \cdot \sum_1^N x_i^2 \right) - \bar{x}^2$$

(průměr čtverců MINUS čtverec průměru měření)

- **Problém: tento odhad je vždy o něco menší než skutečný rozptyl**

druhý adept na odhad neznámého rozptylu:

řešení problému: vždy hodnotu $s^2 = \left(\frac{1}{N} \cdot \sum_1^N x_i^2\right) - \bar{x}^2$
o kousek zvětšíme, a dostaneme už dobrý odhad
neznámého rozptylu σ^2 :

○ Vynásobíme s^2 číslem $\frac{N}{N-1}$, které je větší než 1:

$$\text{est } \sigma^2 = \frac{N}{N-1} \cdot s^2 = \frac{1}{N-1} \cdot \sum_{i=1}^k (x_i - \bar{x})^2$$

Každý odhad rozptylu bude na základě určitého počtu stupňů volnosti:

Úloha: nadiktujte mi 5 reálných čísel
má 5 stupňů volnosti: 3, 20, -345, 6, 18

ALE Úloha: nadiktujte mi 5 reálných čísel, jejichž průměr je
• **10**, má jen 4 stupně volnosti: 3, 20, -345, 6, ... a poslední číslo je už pevně dáno: musí být takové, aby součet daných pěti čísel byl roven 50, pak jejich průměr bude 10

Podobně i při odhadu rozptylu na základě měření N hodnot:

$$\text{est } \sigma^2 = \frac{N}{N-1} \cdot s^2 = \frac{1}{N-1} \cdot \sum_{i=1}^k (x_i - \bar{x})^2$$

... ve vzorci je už vypočten průměr \bar{x} , při výpočtu rozptylu
• určujeme průměr kvadratických odchylek od tohoto průměru, tj. volnost tohoto odhadu je o jednu hodnotu menší, čili $V = N - 1$

Obecně platí pro počty stupňů volnosti v jednom souboru měření:

Počet stupňů volnosti odhadu = počet měření MINUS počet parametrů odhadovaných již dříve

párový *t*-test

(experiment opakovaného měření ... na jednom souboru jedinců provedeme dvojí měření)

Příklad 4:

Chceme zjistit, zda kofein (káva) zvyšuje srdeční tep. Proto u 9 lidí změříme srdeční tep, a potom měření zopakujeme poté, co vypili šálek kávy s kofeinem

Získala se následující data:

Tep bez kofeinu	Tep po šálku kávy	Rozdíly obou měření
70	76	6
60	61	1
49	52	3
72	71	-1
70	81	11
66	70	4
55	55	0
54	61	7
80	89	9

*Ze sloupce odchylek vidíme, že tep se po kávě zvýšil;
chceme ale rozhodnout na základě stat. testu*

K1) $H_0: \mu = 0$ (průměr odchylek je zhruba nulový, srdeční tep nezávisí na kofeinu)

$H_1: \mu \neq 0$ (tep závisí na kofeinu, odchylky mezi oběma situacemi jsou významně různé)

K2) kritérium ... průměr \bar{X} , respektive jeho normovaná hodnota $\frac{\bar{X}-0}{\text{est } \sigma_{\bar{X}}}$... děláme skoro totéž co u veličiny s normálním rozdělením, odečítáme střední hodnotu a dělíme (nikoli odchylkou, ale) odhadem odchylky

K3) $\frac{\bar{x}-0}{\text{est } \sigma_{\bar{x}}}$ nelze popsat normálním rozdělením, protože neznáme rozptyl, ale t-rozdělením, které respektuje větší míru nejasnosti než u normálního rozdělení $N(0;1)$

○ (počet stupňů volnosti odhadu rozptylu je $N-1$)

Vzorec hustoty viz skripta, kap. 11 ... hustota je trochu více „rozevřená“, nabývá významných nenulových hodnot na větším intervalu než jen intervalu $(-3;3)$ u veličiny U

K4) volíme $\alpha = 0,05$, a najdeme kritickou mez neboli interval I pomocí tabulky kritických hodnot

(skripta, kapitola 11)

◦ $t_v (V = 8) = 2,306$

$t_m (V = 8) = -2,306$ (kritické hodnoty jsou symetrické vzhledem k nule)

$$\frac{4,44-0}{1,37} = 3,24 \notin (-2,306; 2,306), \text{ tj. } H_0 \text{ zamítáme!!!!}$$

K5) rozhodnutí testu:

$\bar{X} = 4,44$... průměr daných devíti odchylek

est $\sigma^2 = \overline{s^2} = 17,03$... rozptyl jediné hodnoty měření

est $\sigma^2_{\bar{X}} = \frac{\overline{s^2}}{N} = \frac{17,03}{9} = 1,89$... rozptyl průměru měření!!

est $\sigma_{\bar{X}} = \sqrt{1,89} = 1,37$... odhad odchylky, který potřebujeme

Konstrukce intervalu spolehlivosti pro neznámou střední hodnotu průměrného navýšení srdečního tepu:

To, že tep je ovlivněn užitím kofeinu, vidíme i na začátku našeho testu

- Spíše než provést statistický test bychom potřebovali vědět, o kolik tepů zhruba se zvýší puls užitím dané dávky kofeinu ($\mu_{\bar{x}} \neq 0$... ale čemu se tedy zhruba $\mu_{\bar{x}}$ rovná??)

Interval spolehlivosti tedy podává důležitější informaci než statistický test:

Při konstrukci $(1-\alpha)$ %-ního intervalu spolehlivosti vycházíme z daného statistického testu:

○ H_0 nezamítáme pro $\frac{\bar{x} - \mu_{\bar{x}}}{\text{est } \sigma_{\bar{x}}} \in (-t_k(N-1); t_k(N-1))$

Meze intervalu splňují dvě rovnice, ze kterých vyjádříme $\mu_{\bar{x}}$

Dostaneme :

$$\mu_{\bar{X}} \in (\bar{X} - \text{est } \sigma_{\bar{X}} \cdot t_k(N-1); \bar{X} + \text{est } \sigma_{\bar{X}} \cdot t_k(N-1))$$

A tedy

$$\mu_{\bar{X}} \in (\bar{X} - \sqrt{\frac{s^2}{N}} \cdot t_k(N-1); \bar{X} + \sqrt{\frac{s^2}{N}} \cdot t_k(N-1))$$

○ V našem příkladu $\mu_{\bar{X}} \in 4,44 \pm \sqrt{\frac{17,03}{9}} \cdot 2,306 = (1,27; 7,61)$... puls se s kofeinem zhruba zvýšil o 1 až 7 úderů za minutu

Mezi stat. testem a intervalem spolehlivosti pro totéž α existuje jasný vztah:

$H_0: \mu_{\bar{x}} = \mu_0$ nebude zamítnuta na hladině významnosti α právě tehdy, když $\mu_0 \in$ daného $(1 - \alpha)\%$ -ního intervalu spolehlivosti pro střední hodnotu průměru $\mu_{\bar{x}}$.

○

V našem příkladu $\mu_0 = 0 \notin (1,27; 7,61)$... a to odpovídá i výsledku testu, ve kterém jsme H_0 zamítli

nepárový t-test (typ „dvě skupiny jednou“)

Změříme jednu skupinu za jedné situace (obvyčejně speciální metoda, inovativní podmínka), druhou skupinu za jiné, většinou normální (nijak nepozměněné) situace (tzv. kontrolní skupina)

Příklad 5:

Chceme porovnat dvě techniky zapamatování 100 slov:

- Skupina 1 ... učí se inovativní technikou, $N_1 = 5$ lidí
- Skupina 2 ... učí se klasickou metodou, $N_2 = 7$ lidí

Získala se následující data (počet zapamatovaných slov ze 100):

Metoda inovativní	Metoda klasická
43	16
37	22
51	24
27	30
32	18
//	20
//	27

Provedme statistický test pro tato data:

K1) $H_0: \mu_1 = \mu_2$ (zapamatování nezávisí na technice)

- **$H_1: \mu_1 \neq \mu_2$ (střední hodnoty průměrů obou technik jsou rozdílné, zapamatování závisí na metodě)**

K2) kritérium ... rozdíl průměrů $\bar{X}_1 - \bar{X}_2$, respektive

normovaná hodnota $\frac{\bar{X}_1 - \bar{X}_2 - 0}{\text{est } \sigma_{\bar{X}_1 - \bar{X}_2}}$

K3) $\frac{\bar{X}_1 - \bar{X}_2 = 0}{\text{est } \sigma_{\bar{X}_1 - \bar{X}_2}}$ lze za předpokladu H_0 popsat t-rozdělením, kde odhad rozptylu určíme na základě obou souborů měření (předpokládáme, že vnitřní rozptýlení hodnot je v obou skupinách stejné, tj. že dílčí odhady $\text{est } \sigma^2_1$, $\text{est } \sigma^2_2$ jsou odhady téhož rozptylu ...)

$$\text{est } \sigma^2 = \frac{4}{4+6} \cdot \overline{S_1^2} + \frac{4}{4+6} \cdot \overline{S_2^2} = \frac{4}{4+6} \cdot 88 + \frac{4}{4+6} \cdot 24,61905 = 49,97143$$

(vážený průměr dvou dílčích rozptylů, volnost první = 4, volnost druhá = 6) ... jedná se o rozptyl jednoho měření!!

Nyní platí pro $\text{est } \sigma^2_{\bar{X}_1 - \bar{X}_2}$

$$\text{est } \sigma^2_{\bar{X}_1 - \bar{X}_2} = D(\bar{X}_1) + D(\bar{X}_2) = \frac{\text{est } \sigma^2}{5} + \frac{\text{est } \sigma^2}{7} = \frac{49,97143}{5} + \frac{49,97143}{7} = 17,13306$$

Tedy $\text{est } \sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{17,13306} = 4,13921$

○

$\frac{\bar{X}_1 - \bar{X}_2 = 0}{4,13921}$ lze při platnosti H_0 popsat rozdělením t za $4+6=10$ stupňů volnosti

K4) volíme $\alpha = 0,05$, a najdeme kritickou mez neboli interval I pomocí tabulky kritických hodnot

(skripta viz kap. 11)

○ $t_v (V = 10, \alpha = 20) = 2,228$

$t_m (V = 10, \alpha = 20) = -2,228$

K5) rozhodnutí testu:

$\bar{X}_1 = 38, \bar{X}_2 = 22,42857$... průměry v obou skupinách

$$\circ \frac{\bar{X}_1 - \bar{X}_2 - 0}{4,13921} = \frac{38 - 22,42857 - 0}{4,13921} = 3,761 \notin (-2,228; 2,228),$$

Tj. H_0 zamítáme, inovativní metoda je statisticky významně lepší

Lze sestavit interval spolehlivosti pro každou z obou středních hodnot:

$$\frac{\bar{X}_1 - \mu_1}{\text{est } \sigma_{\bar{X}_1}} \in (-t_k(N_1 - 1 + N_2 - 1); t_k(N_1 - 1 + N_2 - 1))$$

Meze intervalu splňují dvě rovnice, ze kterých vyjádříme μ_1

$$\mu_1 \in (\bar{X}_1 - \sqrt{\frac{49,97143}{5}} \cdot t_k(10); \bar{X}_1 + \sqrt{\frac{49,97143}{5}} \cdot t_k(10)), \text{ a tedy}$$

$$\mu_1 \in 38 \pm \sqrt{\frac{49,97143}{5}} \cdot 2,228 = (38 - 7,043541; 38 + 7,043541) = (30,96; 45,04)$$

Podobně pro μ_2 :

$$\frac{\bar{X}_2 - \mu_2}{\text{est } \sigma_{\bar{X}_2}} \in (-t_k(N_1 - 1 + N_2 - 1); t_k(N_1 - 1 + N_2 - 1))$$

Meze intervalu splňují dvě rovnice, ze kterých vyjádříme μ_2

$$\mu_2 \in (\bar{X}_2 - \sqrt{\frac{49,97143}{7}} \cdot t_k(10); \bar{X}_2 + \sqrt{\frac{49,97143}{7}} \cdot t_k(10)), \text{ a tedy}$$

$$\mu_2 \in 22,42857 \pm \sqrt{\frac{49,97143}{7}} \cdot 2,228 = (22,42857 - 5,95288; 22,42857 + 5,95288) = (16,4757; 28,3814)$$

oba intervaly spolehlivosti pro jednotlivé průměry se vůbec nepřekrývají – to přesně odpovídá tomu, že H_0 v příslušném statistickém testu bylo zamítnuto